ENTERED   PCT09

```
     3 <110> APPLICANT: SHIBATA, Takashi
     4       ICHIKAWA, Chiyo
     5       MATSUURA, Mitsutaka
     6       NOGUCHI, Yuji
     7       SAITO, Yoshimasa
     8       YAMASHITA, Michio
     9       TAKATA, Yoko
    11 <120> TITLE OF INVENTION: SORBITOL DEHYDROGENASE, GENE ENCODING THE SAME AND USE
THEREOF
    13 <130> FILE REFERENCE: 213930US0PCT
    15 <140> CURRENT APPLICATION NUMBER: 09/926,163B
C--> 16 <141> CURRENT FILING DATE: 2002-05-28
    18 <150> PRIOR APPLICATION NUMBER: PCT/JP00/01608
    19 <151> PRIOR FILING DATE: 2000-03-16
    21 <150> PRIOR APPLICATION NUMBER: JP11/72810
    22 <151> PRIOR FILING DATE: 1999-03-17
    24 <150> PRIOR APPLICATION NUMBER: JP11/224679
    25 <151> PRIOR FILING DATE: 1999-08-06
    27 <160> NUMBER OF SEQ ID NOS: 20
    29 <170> SOFTWARE: PatentIn version 3.1
    31 <210> SEQ ID NO: 1
    32 <211> LENGTH: 4115
    33 <212> TYPE: DNA
    34 <213> ORGANISM: Gluconobacter oxydans
    36 <220> FEATURE:
    37 <221> NAME/KEY: CDS
    38 <222> LOCATION: (537)..(1994)
    39 <223> OTHER INFORMATION:
    42 <400> SEQUENCE: 1
    43 aagcttgcat gcctgcaggt cgactctaga ggatccggtt ttggcagcgc tccctagatt    60
    45 gatgcggcgt ctgttgaccg acatgatgct ggtggcacgt gccattgcga cggggcgtgc   120
    47 gaccgggaac acaggcctgc tgcctttgta caaggggctg agtcatgcgc tgcgtgggct   180
    49 ggcacatagt tgcgaagagc agttgcgcgc aaagcagaac cagcatgaac agcagtccga   240
    51 agacgaggaa atcctcggcc tcctaccgcg attggaagag cagacccgtc ctgagatgcg   300
    53 ttttgtgatg tccctgttcc gcgaggatct cgaacgggct gttggggtgc tcatgcgttc   360
    55 tgatgcgagt gccgcaaaag gtctctgaac aggacgtccc gcggagggca gtcagaggtc   420
    57 gaaatggctc ctgttgaaac cgtcattcgg tttttacgtt gtttcggggc tatgatggca   480
    59 catgcccggc cttgtcggtc cccgtcagcg accggcccga aaccacggag aattcc atg   539
    60                                                               Met
    61                                                                 1
    63 att acg cgc gaa acc ctt aag tct ctt cct gcc aat gtc cag gct ccc     587
    64 Ile Thr Arg Glu Thr Leu Lys Ser Leu Pro Ala Asn Val Gln Ala Pro
    65            5                   10                  15
    67 ccc tat gac atc gac ggg atc aag cct ggg atc gtg cat ttc ggt gta     635
```

RAW SEQUENCE LISTING   DATE: 07/24/2002
PATENT APPLICATION: US/09/926,163B   TIME: 11:18:07

Input Set : A:\213930US0PCT.txt
Output Set: N:\CRF3\07242002\I926163B.raw

```
68  Pro Tyr Asp Ile Asp Gly Ile Lys Pro Gly Ile Val His Phe Gly Val
69          20              25                  30
71  ggt aac ttt ttt cga gcc cat gag gcg ttc tac gtc gag cag att ctt       683
72  Gly Asn Phe Phe Arg Ala His Glu Ala Phe Tyr Val Glu Gln Ile Leu
73          35              40                  45
75  gaa cac gct ccg gac tgg gcg att gtt ggt gtt ggc ctg acg ggc agt       731
76  Glu His Ala Pro Asp Trp Ala Ile Val Gly Val Gly Leu Thr Gly Ser
77  50              55                  60                  65
79  gac cgt tca aag aaa aaa gcc gag gaa ttc aag gcc cag gac tgc ctg       779
80  Asp Arg Ser Lys Lys Lys Ala Glu Glu Phe Lys Ala Gln Asp Cys Leu
81              70                  75                  80
83  tat tcc ctg acc gag acg gct ccg tcc ggc aag agc acg gtg cgc gtc      827
84  Tyr Ser Leu Thr Glu Thr Ala Pro Ser Gly Lys Ser Thr Val Arg Val
85              85                  90                  95
87  atg ggc gcg ctg cgt gac tat ctg ctt gcc ccg gcc gat ccg gaa gcc      875
88  Met Gly Ala Leu Arg Asp Tyr Leu Leu Ala Pro Ala Asp Pro Glu Ala
89          100             105                 110
91  gtg ctg aag cat ctt gtt gat ccg gcc atc cgc atc gtt tcc atg acg      923
92  Val Leu Lys His Leu Val Asp Pro Ala Ile Arg Ile Val Ser Met Thr
93      115             120                 125
95  atc acg gaa ggc ggc tac aac atc aac gag acg acc ggt gcg ttc gat      971
96  Ile Thr Glu Gly Gly Tyr Asn Ile Asn Glu Thr Thr Gly Ala Phe Asp
97  130             135                 140                 145
99  ctg gag aat gcg gca gta aag gcc gac ctc aag aac ccg gaa aag ccg     1019
100 Leu Glu Asn Ala Ala Val Lys Ala Asp Leu Lys Asn Pro Glu Lys Pro
101             150                 155                 160
103 tct acc gtt ttc ggt tac gtg gtc gag gcc ctg cgt cgt cgt tgg gat    1067
104 Ser Thr Val Phe Gly Tyr Val Val Glu Ala Leu Arg Arg Arg Trp Asp
105             165                 170                 175
107 gcc ggt ggt aag gca ttt acg gtc atg tcc tgt gat aac ctg cgt cat    1115
108 Ala Gly Gly Lys Ala Phe Thr Val Met Ser Cys Asp Asn Leu Arg His
109         180                 185                 190
111 aac ggc aat gtc gcc cgc aag gcc ttc ctc ggc tat gcg aag gcg cgc    1163
112 Asn Gly Asn Val Ala Arg Lys Ala Phe Leu Gly Tyr Ala Lys Ala Arg
113     195                 200                 205
115 gat ccg gag ttg gcg aag tgg att gag gaa aac gcg acc ttc ccg aac    1211
116 Asp Pro Glu Leu Ala Lys Trp Ile Glu Glu Asn Ala Thr Phe Pro Asn
117 210                 215                 220                 225
119 gga atg gtt gat cgc atc acc ccg acc gtt tcg gcg gaa atc gcc aag    1259
120 Gly Met Val Asp Arg Ile Thr Pro Thr Val Ser Ala Glu Ile Ala Lys
121             230                 235                 240
123 aag ctc aac gcg gcc agt ggg ctg gat gac gac ctg ccg ctg gtg gcc    1307
124 Lys Leu Asn Ala Ala Ser Gly Leu Asp Asp Asp Leu Pro Leu Val Ala
125             245                 250                 255
127 gag gat ttc cat cag tgg gtg ctg gaa gac cag ttt gcg gat ggc cgt    1355
128 Glu Asp Phe His Gln Trp Val Leu Glu Asp Gln Phe Ala Asp Gly Arg
129         260                 265                 270
131 ccg ccg ctt gaa aaa gcc ggc gtg cag atg gtc ggg gac gtg acg gac    1403
132 Pro Pro Leu Glu Lys Ala Gly Val Gln Met Val Gly Asp Val Thr Asp
```

RAW SEQUENCE LISTING   DATE: 07/24/2002
PATENT APPLICATION: US/09/926,163B   TIME: 11:18:07

Input Set : A:\213930US0PCT.txt
Output Set: N:\CRF3\07242002\I926163B.raw

```
133     275                    280                    285
135 tgg gag tac gtc aag atc cga atg ctc aat gca ggg cat gtc atg ctc    1451
136 Trp Glu Tyr Val Lys Ile Arg Met Leu Asn Ala Gly His Val Met Leu
137 290                    295                    300                    305
139 tgc ttc cca ggc att ctg gtc ggc tat gag aat gtg gat gac gcc att    1499
140 Cys Phe Pro Gly Ile Leu Val Gly Tyr Glu Asn Val Asp Asp Ala Ile
141                    310                    315                    320
143 gaa gac agc gaa ctc ctt ggc aat ctg aag aac tat ctc aac aag gat    1547
144 Glu Asp Ser Glu Leu Leu Gly Asn Leu Lys Asn Tyr Leu Asn Lys Asp
145                    325                    330                    335
147 gtc atc ccg acc ctg aag gcg cct tca ggc atg acg ctc gaa ggc tat    1595
148 Val Ile Pro Thr Leu Lys Ala Pro Ser Gly Met Thr Leu Glu Gly Tyr
149         340                    345                    350
151 cgg gac agc gtc atc agc cgt ttc tcc aac aag gcg atg tcg gac cag    1643
152 Arg Asp Ser Val Ile Ser Arg Phe Ser Asn Lys Ala Met Ser Asp Gln
153     355                    360                    365
155 acg ctc cgg att gct agc gat ggc tgt tcc aag gtt cag gtg ttc tgg    1691
156 Thr Leu Arg Ile Ala Ser Asp Gly Cys Ser Lys Val Gln Val Phe Trp
157 370                    375                    380                    385
159 acg gaa acc gtg cgt cgg gcg atc gaa gac aag cgg gac ctg tca cgt    1739
160 Thr Glu Thr Val Arg Arg Ala Ile Glu Asp Lys Arg Asp Leu Ser Arg
161                    390                    395                    400
163 ata gcg ttc gga att gca tcc tat ctc gaa atg ctg cgt ggt cgc gac    1787
164 Ile Ala Phe Gly Ile Ala Ser Tyr Leu Glu Met Leu Arg Gly Arg Asp
165                    405   .                410                    415      .
167 gag aag ggc ggg acg tat gaa tcg tcc gag ccg act tat ggc gac gcc    1835
168 Glu Lys Gly Gly Thr Tyr Glu Ser Ser Glu Pro Thr Tyr Gly Asp Ala
169         420                    425                    430
171 gaa tgg aag ttg gcc aag gcg gac gac ttc gaa agc tct ctg aag ctc    1883
172 Glu Trp Lys Leu Ala Lys Ala Asp Asp Phe Glu Ser Ser Leu Lys Leu
173     435                    440                    445
175 ccg gcg ttc gat ggg tgg cgc gat ctg gat acg tcc gaa ctg gat caa    1931
176 Pro Ala Phe Asp Gly Trp Arg Asp Leu Asp Thr Ser Glu Leu Asp Gln
177 450                    455                    460                    465
179 aag gtc atc gtg ctg cgg aag atc atc cgc gaa aag ggc gta aaa gcc    1979
180 Lys Val Ile Val Leu Arg Lys Ile Ile Arg Glu Lys Gly Val Lys Ala
181         470                    475                    480
183 gcc atc ccg gcc tga attcggcttt tagggtagcg actgaaacag aaaaccgcgc    2034
184 Ala Ile Pro Ala
185                    485
187 tctggaagga gcgcggtttt ttttatgctc agatctgtcc catcaggaca aggatcacga    2094
189 cgaccacgat caggacaagt ccgctggagg gggagcccca tttcgaactg tacggccatg    2154
191 acggcagcgc accgagatca ggattacaag aaggatcagt cccatggcac atctctcttg    2214
193 ccggttgaga ctggtctgtg ttccgggtgt ctaaaaagtt ccgtagggg cgcgaaagat    2274
195 caaagctgtc ggtcgcgctt aatccggtcc caagccgcat tgatgcgggc cacccggtcc    2334
197 tgtgcgcgtt tgcgctctgt ctctgacata ggtttctggg ccagcacgtc cggatgatgt    2394
199 tcgcggatca gggtgcgcca gcgcacgcgg atttctgtgt cagttgcgct gcgggtgatg    2454
201 ccgagaatac gataggcatc cggctcgttt ccgctggcgg cgcgattgtt ccgctttcg    2514
203 gcccggtccc atgctcctgg cggcaggcca aatgccccgt gaacgcgctg cagaaaatcg    2574
```

```
205 atttccttcg ggtgaagctc gcggctgggg ccggcatcgg cacgggcgat acggaacagt 2634
207 gccgtcatga ggttctcaag cggcgccgta ttatcggcat aggccttgcc catttcgcgg 2694
209 gcatacatct cgaaatcgtc cgtccggtcg cgggcgcgat cgaacagcat gccgacttcc 2754
211 ttggtgttat cggggggggaa ctggaagcag gtcttgaaag cgttgatttc gtgtcggttc 2814
213 accggcccgt cgatcttcgc cagcttcgcg cacagggcaa caaggccgat ggcgtaaagc 2874
215 tgatctcgtt tgcccagggc cgcagcaatc ttggcagcgc cgaaaaaggc cgcgctgttg 2934
217 ggatcgggac ggccattcgc gggaaagcgc tcactccagc cgcccgttga gggcttgagt 2994
219 agcgaaccgt tatcggcggc atgccccagc gctgcgccca tcagtgctcc gaaaggacca 3054
221 ccaaccgcga agcccgcgac accaccgaac atcttgcccc agatagccat gtcatcaacc 3114
223 tagcacgccc gctcacagcg gcaaatgaca gatcgcaggc taggtgtagg tgctgatgcg 3174
225 ccaaccgccc gggcttgcgg tgtggtagaa gctaggagtt acgaacttat cgctgtctca 3234
227 tgcttttgag gcgcaggttc ttctgttcgt ttcatgacgg atatttttat gcccaccttg 3294
229 atccagactg ctacttcgat ccctttccgc tctgatgacg aactgatgga tcttttgatc 3354
231 aagcgtctgc caatgtggct gcagaaagtg ctgaactggt tgcgggaagc ggatcataaa 3414
233 tgggttcgga ttccggcggg cgtgctgttc atgctgggcg gcgttctgtc catcctgcct 3474
235 gttctgggtc tgtggatgct gccggtcggc gtgatgttgc ttgcgcagga tattccgttc 3534
237 ttccgtcgcc ttcagggccg cctcttgcgc tggatcgaac gtcaacatcc ggattggctg 3594
239 ggccttccgg cgaaaagcgg cagaagctaa ccgttcgtct ggacgtgttt ctgaagatgt 3654
241 gtcagtgctg caacccgcag ggctgaagcc agtgggcgct ctggtggtcg cgcggcatcg 3714
243 agagaagcca ccagagacgc aaagctctgc tggcggactg cggccatcgc gtccagtata 3774
245 gcccagaact cgggttccag tgccacggac gtccggtgtc ctgacagaga caggctgcgt 3834
247 ttgacgagat cactcattcc ggttgtttct caaggcgctt caaagcccat tgtgcggttt 3894
249 cggaaacatc agggtccgga tcactcagca gctcccgcgc agaagatata agcgacggat 3954
251 cggccgagtt gccgatcgcg atcaggacag ttacgtacga accggttgcg tccaatccgt 4014
253 ttgaccggag agccagaaaa aaacgtccgg aatgtcgcat tatccagccg caccagttcg 4074
255 tcgagttttg gtgcaatcag ctccgggcgg gcctgaagct t 4115
258 <210> SEQ ID NO: 2
259 <211> LENGTH: 485
260 <212> TYPE: PRT
261 <213> ORGANISM: Gluconobacter oxydans
263 <400> SEQUENCE: 2
265 Met Ile Thr Arg Glu Thr Leu Lys Ser Leu Pro Ala Asn Val Gln Ala
266 1             5                  10                 15
269 Pro Pro Tyr Asp Ile Asp Gly Ile Lys Pro Gly Ile Val His Phe Gly
270            20                 25                 30
273 Val Gly Asn Phe Phe Arg Ala His Glu Ala Phe Tyr Val Glu Gln Ile
274         35                 40                 45
277 Leu Glu His Ala Pro Asp Trp Ala Ile Val Gly Val Gly Leu Thr Gly
278    50                 55                 60
281 Ser Asp Arg Ser Lys Lys Lys Ala Glu Glu Phe Lys Ala Gln Asp Cys
282 65                 70                 75                 80
285 Leu Tyr Ser Leu Thr Glu Thr Ala Pro Ser Gly Lys Ser Thr Val Arg
286            85                 90                 95
289 Val Met Gly Ala Leu Arg Asp Tyr Leu Leu Ala Pro Ala Asp Pro Glu
290         100                105                110
293 Ala Val Leu Lys His Leu Val Asp Pro Ala Ile Arg Ile Val Ser Met
294         115                120                125
297 Thr Ile Thr Glu Gly Gly Tyr Asn Ile Asn Glu Thr Thr Gly Ala Phe
298    130                135                140
```

RAW SEQUENCE LISTING   DATE: 07/24/2002
PATENT APPLICATION: US/09/926,163B   TIME: 11:18:07

Input Set : A:\213930US0PCT.txt
Output Set: N:\CRF3\07242002\I926163B.raw

```
301 Asp Leu Glu Asn Ala Ala Val Lys Ala Asp Leu Lys Asn Pro Glu Lys
302 145             150             155             160
305 Pro Ser Thr Val Phe Gly Tyr Val Val Glu Ala Leu Arg Arg Arg Trp
306             165             170             175
309 Asp Ala Gly Gly Lys Ala Phe Thr Val Met Ser Cys Asp Asn Leu Arg
310             180             185             190
313 His Asn Gly Asn Val Ala Arg Lys Ala Phe Leu Gly Tyr Ala Lys Ala
314         195             200             205
317 Arg Asp Pro Glu Leu Ala Lys Trp Ile Glu Glu Asn Ala Thr Phe Pro
318     210             215             220
321 Asn Gly Met Val Asp Arg Ile Thr Pro Thr Val Ser Ala Glu Ile Ala
322 225             230             235             240
325 Lys Lys Leu Asn Ala Ala Ser Gly Leu Asp Asp Asp Leu Pro Leu Val
326             245             250             255
329 Ala Glu Asp Phe His Gln Trp Val Leu Glu Asp Gln Phe Ala Asp Gly
330             260             265             270
333 Arg Pro Pro Leu Glu Lys Ala Gly Val Gln Met Val Gly Asp Val Thr
334         275             280             285
337 Asp Trp Glu Tyr Val Lys Ile Arg Met Leu Asn Ala Gly His Val Met
338     290             295             300
341 Leu Cys Phe Pro Gly Ile Leu Val Gly Tyr Glu Asn Val Asp Asp Ala
342 305             310             315             320
345 Ile Glu Asp Ser Glu Leu Leu Gly Asn Leu Lys Asn Tyr Leu Asn Lys
346             325             330             335
349 Asp Val Ile Pro Thr Leu Lys Ala Pro Ser Gly Met Thr Leu Glu Gly
350             340             345             350
353 Tyr Arg Asp Ser Val Ile Ser Arg Phe Ser Asn Lys Ala Met Ser Asp
354         355             360             365
357 Gln Thr Leu Arg Ile Ala Ser Asp Gly Cys Ser Lys Val Gln Val Phe
358     370             375             380
361 Trp Thr Glu Thr Val Arg Arg Ala Ile Glu Asp Lys Arg Asp Leu Ser
362 385             390             395             400
365 Arg Ile Ala Phe Gly Ile Ala Ser Tyr Leu Glu Met Leu Arg Gly Arg
366             405             410             415
369 Asp Glu Lys Gly Gly Thr Tyr Glu Ser Ser Glu Pro Thr Tyr Gly Asp
370             420             425             430
373 Ala Glu Trp Lys Leu Ala Lys Ala Asp Asp Phe Glu Ser Ser Leu Lys
374         435             440             445
377 Leu Pro Ala Phe Asp Gly Trp Arg Asp Leu Asp Thr Ser Glu Leu Asp
378     450             455             460
381 Gln Lys Val Ile Val Leu Arg Lys Ile Ile Arg Glu Lys Gly Val Lys
382 465             470             475             480
385 Ala Ala Ile Pro Ala
386             485
389 <210> SEQ ID NO: 3
390 <211> LENGTH: 16
391 <212> TYPE: DNA
392 <213> ORGANISM: Artificial Sequence
394 <220> FEATURE:
```

VERIFICATION SUMMARY   DATE: 07/24/2002
PATENT APPLICATION: **US/09/926,163B**   TIME: 11:18:08

Input Set : **A:\213930US0PCT.txt**
Output Set: **N:\CRF3\07242002\I926163B.raw**

L:16 M:271 C: Current Filing Date differs, Replaced Current Filing Date